File Import Settings

Contents:

- *Per-file encoding*
- *Detect structure*
- *Remove special characters from column names*
- *Selecting column headers*

When you edit settings on a selected file in the Import Data page, the following settings are displayed.

You can edit any additional or optional settings for an individual dataset. Perform the following:

1. Click **Edit Settings** from the card for an individual dataset in the right panel. The dialog box is displayed.
2. In the dialog box, select the required options and modify the settings.

Per-file encoding

By default, Trifacta Wrangler Pro applies a specified encoding type on the imported the file. In some cases, the data preview panel may contain garbled data, due to a mismatch in encodings. In the Data Preview dialog, you can select a different encoding for the file. When the correct encoding is selected, the preview displays the data as expected.

> **NOTE:** Assessing the file encoding type based on parsing an input file is not an accurate method. Instead, Trifacta Wrangler Pro assumes that the file is encoded in the default encoding. If it is not, you should change the encoding type for the file.

> **NOTE:** In some cases, imported files are not properly parsed due to issues with encryption types or encryption keys in the source datastore. For more information, please contact your datastore administrator.

For a list of supported encoding types, see *Supported File Encoding Types*.

Detect structure

By default, Trifacta Wrangler Pro attempts to interpret the structure of your data during import. This structuring attempts to apply an initial tabular structure to the dataset.

- Unless you have specific problems with the initial structure, you should leave the Detect structure setting enabled. Recipes created from these imported datasets automatically include the structuring as the first, hidden steps. These steps are not available for editing, although you can remove them through the Recipe panel. See *Recipe Panel*.
- When detecting structure is disabled, imported datasets whose schema has not been detected are labeled, **unstructured datasets**. When recipes are created for these unstructured datasets, the structuring steps are added into the recipe and can be edited as needed.
- For more information, see *Initial Parsing Steps*.

Remove special characters from column names

When selected, characters that are not alphanumeric or underscores are stripped, and space characters are converted to underscores.

For more information, see *Sanitize Column Names*.

Selecting column headers

You can apply the column headers to your datasets during import. Select the required option from the drop-down list:

- **Infer header**: (default) When selected, the Trifacta application infers the header based on the data in the import.
- **Use first row as header**: When selected, the first row is used as the column headers.
- **No header**: When selected, the inference is ignored and column headers are defined using generic names with no headers.

If replacing a file:

- If you replace a dataset in a flow and select the **Use first row as header** option, then the existing header row labels are updated with the new headers.
- Subsequent steps in a pre-existing recipe may be broken if the headers are changed by a replaced file.

> **Tip:** After the dataset is imported, you can rename columns manually or using any row in the dataset. For more information, see *Rename Columns*.